DOCUMENT RESUME

ED 266 465 CS 209 552

AUTHOR	Atwood, N.; And Others
TITLE	Cognitive Approaches to the Analysis of Technical Understanding. CDC Technical Report No. 24.
INSTITUTION	Carnegie-Mellon Univ., Pittsburgh, PA. Communications Design Center.
REPORT NO	CDC-TR-24
PUB DATE	Nov 85
NOTE	54p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Abstract Reasoning; Comparative Analysis; Content Analysis; Employment Experience; *Paraprofessional Personnel; *Prior Learning; *Readability; *Reading Comprehension; *Reading Research; Technical Writing
IDENTIFIERS	*Text Structure

ABSTRACT

Eighteen naval fire technicians served as subjects in a study using scenarios to simulate a real world task for assessing the readability of general technical texts. The scenario asked subjects to read a segment of text and to explain to a colleague how the particular component described in the passage functioned. Subjects' summaries were assessed on the extent to which both the gist of the passage and specific information were represented. Two experimental variations examined their effects on subjects' summaries in response to the scenario. The first variation involved the text with an original version and a revised version that was simplified using traditional readability indices. The second variation centered on the world knowledge that subjects brought to the task, with three groups of students differing in their amount and type of experience. Results supported the initial distinctions drawn between text versions and experience groups. The revised version was syntactically simpler on a number of dimensions such as readibility level, sentence length, and sentence structure. Furthermore, while the overall semantic content or gist of corresponding segments remained comparable, the revised version was lower in density of ideas, while the ideas presented were higher in hierarchical level and salience than in the original version. Three pages of references conclude the paper. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED266465

CDC Technical Report No. 24

Cognitive Approaches to the Analysis of Technical Understanding

N. Atwood, E. Baker, and T.M. Duffy

November 1985

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Thomas Duffy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS209552

The Communications Design Center was established in 1979 to solve communication problems in business, industry, government, and the professions.

The goals of the Center are:

- to encourage interdisciplinary approaches to the solution of visual and verbal communication problems;
- to promote cooperation between communication researchers at the University and those outside the University who have practical interests in communication problems;
- to provide resources and an organizational setting for research in visual and verbal communication;
- to train professional writers; researchers, and teachers;
- to promote the development of educational materials for communication programs on and off the CMU campus.

This report is one of a series from our on-going research on communication and document design. If you have questions or comments on this report, or if you would like to find out more about the activities of the CDC, contact:

Director
Communications Design Center
Baker Hall 160
Carnegie-Mellon University
Pittsburgh, PA 15213
(412) 578-2906

There was, and perhaps still is, a belief that computerization will lead to the "paperless" society and thus to a dramatic reduction in the requirement for reading. However, as virtually anyone who has entered the age of computers can attest, that prediction thus far has proved to be a wild fantasy. Not only isn't there a decrease in the volume of paper, there is a significant increase both in the paper AND in the requirement for reading. Indeed, as we move from a manufacturing based economy to an economy based on the delivery of services, we see an increase in the amount and the variety of reading demanded. Thus the goal underlying the initial development of readability formulas, matching readers to test, is even more important today. We must ensure that the text is a useful tool, that is, that people can obtain the information which is necessary to carry out professional and other day to day activities. We need a mechanism for matching a text to an audience -- not only in the selection of a text for an audience but also in guiding the initial design of the text.

Readability formulas and the accompanying readable writing guidelines were designed to provide just such a matching of text to audience (Klare 1963, 1975; Flesch 1948, 1949). A "formula" is actually a regression equation used to predict comprehension. Thus, characteristics of the text are measured and entered into a regression equation to determine the "best" predictor function. In using the formula it is only necessary to measure the predictor variables (word and sentence length factors for most formulas) and enter them into the calculations to obtain the comprehension measure (typically a reading skill requirement). Clearly, a readability formula so developed is an empirical tool. There is complete agreement

that the predictor variables are empirically determined and thus do not necessarily represent causative factors.

While readability formulas are empirical devices they are based on some model of the reading process (Duffy, 1985a). For example, all of the decisions involved in developing and applying a formula are based on some model of the reading process. Predictors are sampled from the set of variables thought to be causally related to comprehension (e.g., word familiarity) or to serve as proxy measures for such variables (e.g., word length). The universe sampled (and by implication the universe not sampled) reflects concepts of the determinants of comprehension. The test used to assess comprehension of the passages provides a definition of what is meant by comprehension. And the index of skill requirements, what the formula predicts, provides an indication of the skill requirements for reading.

There is also a model of the reading process implicit in the application of readability formulas (Duffy, 1985a). For example, the extent to which one is willing to apply a given formula to different audiences, different varieties of texts, and different reading tasks indicates the degree to which reading is viewed as a homogeneous process. When a text is identified as "too difficult", the variables in the formula indicate the kind of simplication which must occur, that is, the variables which must be attended to in the revision process. While the text is never written to the formula, the simplification must be based on the class of variables underlying the predictors.

There are well over 100 readablility formulas in popular use today (Klare, 1979) and all of those formulas are based on the same underlying

model of the reading process. It is the limitations and faulty assumptions of that model which are the basis of most of the criticism of readability formulas. The model is basically the traditional model of reading which is the basis of most reading instruction. In this model reading is seen as a relatively passive activity in which the reader decodes the text to obtain the meaning (Duffy, in press). Thus it is a bottom up model in common parlance. The decoding is at the level of sentences and of words. Thus "reading" can be defined in terms of the skills necessary to decode words and sentences. These skills are not dependent on content knowledge except as reflected in the knowledge of specific vocabulary. Thus skill training is independent of the content domain of the text; instruction is, in general, directed toward reading ability. Furthermore the level of skill, i.e., the level of general reading ability, applies equally in all content domains.

The application of this model to the analysis of text leads to two primary effects. First, since reading is viewed as decoding words and sentences, the difficulty of the text must be indexed in terms of word and sentence characteristics. Thus even the most complex of the popular readability formulas (e.g., Bormuth, 1979) sampled only syntactic and lexical features to identify the "best" predictors. Second, since reading is viewed as a general process independent of domain knowledge, the formulas have wide applicability. The formulas are applied regardless of the comprehension task (e.g., job related reading, learning in a class, finding facts), regardless of the subject matter expertise of the audience, and regardless of the nature of the text (narrative, expository, technical) (Duffy, 1985a).

BEST COPY AVAILABLE

Our understanding of the reading process has grown considerably beyond the simple decoding model. Reading is not simply finding the meaning in the text, but rather meaning is generated based on the text representation and world knowledge. The reader draws from long term memory to build a a coherent structure. Understanding is based on the information goals of the reader, his or her knowledge of the subject matter, and the representation of the information in the text. Thus there is a top down imposition or construction of meaning along with the bottom up deoding of the text (Wittrock, 1982; Spiro, Bruce, and Brewer, 1980).

The importance of this top down process has been illustrated in numerous studies. Perhaps the classic example is a study by Bransford and Johnson (1972) in which a paragraph concerning moving things about made no sense and was very difficult to remember until a title, washing clothes, was provided. The title provided a link to prior knowledge; it provided access to a schema or structure for interpreting activities involving the washing of clothes. The text remained the same, yet there was dramatic change in comprehension. Clearly, this was a result that the readability formula score could not predict. Indeed, how would one calculate the readability?

The use of prior knowledge in the form of schemas for organizing and interpreting text is a major construct in current theories of reading. Kieras (1985) has argued that, in addition to subject domain or content relevant schemas, readers employ text structure schemas to interpret text. Content schemas are organizational structures for a content domain -- an organization of slots in which facts are to be inserted. The text schema is an organizational structure for texts independent of content. He

BEST COPY AVAILABLE

proposes that there are common text schemas which specify where in a passage specific content is likely to appear. These text schemas are especially important in the comprehension of technical text where the reader frequently has limited knowledge of the content domain.

Text schemas basically reflect that readers learn how to "use" a text. If a new text has the same structure, their text schema will yield positive transfer. However, if a new presentation involves a new structure, the negative transfer of text schemas could well counteract any content or readability based on improvements in the text. The effects of this text schema have been hypothesized to account for the failure of several major revision efforts to improve the usability of a technical manual (Duffy, Curran, and Sass, 1983).

Pichert and Anderson (1977) have demonstated the importance of the goal in reading on the contents and organization of memory. In one study, readers read the description of a house either from the perspective of a burgler or from the perspective of a home buyer. The characteristics of the house that were retained was a function of the particular perspective of the reader. The burglar organized recall and remembered things appropriate to planning a burglary while the home buyer remembered features of the house. While not a part of the Pichert and Anderson study, the paragraph could have been written to support one or the other interpretation (see, e.g., Frase, 1977). Thus, the readability of the text would depend on the degree to which the organizational structure of the text was consistent with the information goals of the reader. This issue has been of critical concern in the design of technical text as reflected in the discussion of functional vs. topical organizational structures

(Duffy, 1985a). The topic organization may be adequate if the comprehension task is learning about a particular component but a functional organization is essential to using the text to work on the equipment.

Osborne and Wittrock (1983) have demonstrated that.....Note to editor: paragraph forthcoming).

These studies demonstrate the active process of meaning generation involved in reading. Prior knowledge, the goal in reading, and the interrelationship of text units all play a critical role in comprehension. Traditional readability formulas do not account for these factors in the reading process. As a consequence, the formulas are theoretically inappropriate and uninteresting. At a more practical level, they simply provide an inadequate prediction of comprehension and inadequate guidance for revision. In fact, the revision guidance may simply be incorrect. For example, simplifying sentences with the intent of improving comprehension may in fact reduce comprehension through an increased requirement for anaphora and an increased separation of the anaphora and referent in the processing of the text (Pearson, 1974, 1975; Duffy and Kabance, 1982).

The inadequacy of the predictions arising from traditional readability formulas is not simply a matter of providing an increment in the precision of prediction. Rather new approaches are required to account for dramatic shifts in the usability of a text when the organization of the sentences, the subject matter knowledge of the user, or the comprehension task is changed. Users of traditional formulas may well be rejecting texts which an audience of subject matter experts could readily use or accepting texts where the audience of new learners will have considerble difficulty.

There have been several recent attempts to develop text analysis schemes more appropriate to the active processing model of reading. The best known of these are the approaches of Miller and Kintsch (1980) and Meyer (1975). In the Miller and Kintsch model, text is analyzed into propositions and it is these propositions which are used to build a meaning structure. An initial text structure or schema is imposed based on the initial propositions. That structure is expanded, modified, or abandoned as the reader attempts to interpret successive propositions in terms of it. Frequency of occurrence of propositions and the limits of short term memory in holding sequences of propositions play important roles in determining comprehensibility.

Meyer's (1975) approach is also based on a propositional analysis of the text and the identification of the coherence relations between the propositions. From her theoretical perspective, the reader builds a hierarchical structure of the propositions. Propositions higher in the structure or supporting the higher order propositions will be better recalled. Meyer's analysis identifies the structure imposed by the author. However, she recognizes that readers may or may not adopt that structure. Failure to adopt the authors' structure may be due to inadequate signalling of the structure by the author, lack of skill on the part of the reader, or a difference in goals (information requirements) between the reader and writer.

While the Meyer and Miller and Kintsch models account for semantic relations between sentences and address the reader's limitations in processing, the analyses are independent of the reader's knowledge and

goals. Thus they fail, as it would seem any text analysis system must fail, to provide a generalizable methodology for adequately indexing the readability of a text for a particular audience. Miller and Kintsch (1980) propose that their model recognizes readability as "an interction between the text and the reader's prose processing capabilities, rather than as some innate property of text." However, traditional readability approaches also presume an active decoder with prose processing capabilities. The Miller and Kintsch model simply extends beyond sentence boundaries and includes semantic relations in the text. However, neither model identifies differential processing strategies as a function of different reading goals. All text is assumed to be processed in the same way. Furthermore, processing is presumed to proceed sentence by sentence or proposition by proposition from the beginning of the text to the end. Finally, while Miller and Kintsch identify world knowledge as an important variable in comprehension, they have not yet incorporated it into their analysis approach. In essence the approach cannot account for the findings of Bransford and Johnson (1972), Pichert and Anderson (1977), or Osborne and Wittrock (1983); thus, readability does in fact remain inherent in the text.

Indeed any text analysis scheme that excludes the reader and the reading task will by necessity provide an inadequate account of the reading process and inadequate predictions of the usability of the text. Reading in the real world is not the free recall of 250 or 2,000 word passages nor is it the ability to fill in words deleted from a text. It is mysterious to us why one would suppose that the processes or strategies in carrying

out these types of tasks should at all resemble the process and stategies of using text in real world situations. If the processes differ, then we may well expect the relevant text features to differ.

There is sufficient evidence that assessment of the ability of the audience to use the text and and assessment of how they use the text in some meaningful task, provides valuable information as to both the reading process and the assessment of readability. Indeed such an analysis process has become a standard in problem solving research (Erickson and Simon, 1984) as a means of understanding the problem solving process. At the practical level, it would also seem that actual testing of the ability and processes of specific audiences to use text in a meaningful way is essential to effective prediction.

The alternative to actual testing is either the use of prediction formulas or the application of the judgment of expert writers. We have already discussed the inadequacies of prediction formulas. The judgment of expert writers does not fair much better. Carver (1974) and Wright (1985) have found that experienced writers cannot predict the readability of text. Swayney, Janick, Bond, and Hayes (1981) had expert writers revise four different consumer contracts to make them more usable by the intended audience. Comprehension was then assessed by asking representatives of that audience to use a contract to answer specific application questions. That is, scenarios of ways in which the contract would actually be used were created. The results indicated that the expert revision resulted in improved "readability" of one document, decreased "readability" of another, and produced no change in the other two. However, when the writers had the

BEST COPY AVAILABLE

data available from user testing, including transcripts of the users verbalizations (they were instructed to think out loud while they answered the questions), they were able to significantly improve the readability of the documents.

It has long been a contention that text must be evaluated through testing with a segment of the audience (Redish, Felke, and Rase, 1981). Readability approaches were developed primarily as an attempt to side step the effort involved in such testing, though it also represents an approach to cataloging relevant comprehension factors (e.g., Amarin and Jones, 1982). However, the approach has failed to adequately represent the comprehension tasks, and thus the relevant processing variables, as well as to capture the important role of world knowledge in interpreting text. Thus we would argue that a more fruitful approach to both prediction and understanding would be to develop effective strategies for audience testing using real scenarios. This approach has become standard in the evaluation of procedural information at the Communications Design Center at Carnegie-Mellon University. When computer manuals are written, the procedural information is tested by asking users to think outloud while performing the task described in the document. An analysis of the user's strategies for interpreting and using the text is fed back into the revision process.

In the remainder of this chapter we present data on the use of scenarios to assess the readability of more general, technical text. The text presented the theory of operation (functional description) of a radar receiver. The comprehension goal in using this type of text is not to

learn and remember as much as possible. Rather it seems to have three uses: 1) as simply a reference point for facts about the system; 2) as a source for building a representation the the system; and 3) as a means of providing more detailed information on the functioning of a component. We focused on assessing the latter two comprehension goals. Thus we asked subjects to read a segment of text and then to use that information to describe to a colleague how the particular component functioned. We measured the extent to which the subject summarized in terms of key information units and the gist of the segment. We assessed the effects of the text on subjects' development of a functional representation of the system by asking them to provide a graphic representation of the functional relationships. The graphic representation was collected as a pretest and a postest.

The research also looked at the effects of traditional simplification strategies and of world knowledge on these comprehension measures. We used two versions of the text: an original and a revision based on meeting a variety of readability indices. World knowledge was manipulated by varying the expertise of the users. Expertise was defined in terms of both system knowledge and knowledge of other systems.

Thus, while the research has only limited scope and generalizability, it was designed to provide information related to the following research questions:

1. What were the syntactic and semantic characteristics of the original and revised versions of technical text?
2. How did the original and revised versions of technical text influence the verbal summaries of subjects?

3. What was the effect of different levels of world knowledge on subjects' summaries?

4. What differences were there in subjects' summaries of main ideas and specific information?

5. How did text version and world knowledge influence subject summaries of main ideas and specific information?

6. What was the relationship between subjects' graphic representations of the system and text version, world knowledge, and verbal summaries?

STUDY PROCEDURES

Subjects

Eighteen naval fire technicians served as subjects in the experiment. Subjects were selected to represent three groups of technicians who differed in their world knowledge as indicated by the amount and nature of their experience: (1) expert technicians who were trained and assigned to the electronic system containing the radar receiver described by the technical text (the NATO Seasparrow system*); (2) novice technicians who were not trained in school on the NATO Seasparrow system (NSS) but were assigned to the system for on the job training; and (3) expert technicians who were trained and assigned to another system other than the NATO Seasparrow system.

Subjects were selected from ships in port at a major naval seaport during the two month period in which data collection for the study was conducted. Technicians were identified as experts or novices on the

*NATO refers to the North Atlantic Treaty Organization. Member countries jointly supported the development of the Seasparrow system.

Seasparrow or another system based on the report of their commanding officer. Their classification was verified by inspecting Navy personnel records to determine whether they had attended school training on the NATO Seasparrow system, the length of their on-the-job training, and the specific system to which they were assigned. The personnel records verified the commanding officer classifications.

Personnel records indicated that six of the subjects had attended NATO Seasparrow School and had an average of 24 months of on-the-job training on the NATO Seasparrow system ("Experienced NSS Technicians"). Six subjects had not attended NATO Seasparrow School and had an average of 15 months of on-the-job training on the NATO Seasparrow system ("Novice NSS Technicians"). The remaining six subjects had not attended NATO Seasparrow School, had no on-the-job training on the NATO Seasparrow system but had received formal school training and on-the-job training on another system ("Experienced non-NSS Technicians").

Text Versions

Two versions of a section of technical text were used in the study. The text was drawn from a section of the technical manual for the NATO Seasparrow system that is intended to provide a general functional description of the system. One part of this section, that describing the air designate mode of the radar receiver system was selected for use in the study. This part was selected because it was judged as representative of the type and range of information typically presented in functional system descriptions by a naval subject matter expert.

The text was divided into 25 segments judged to reflect discrete sets of information. This segmentation was made to facilitate the experimental process so that protocols could be collected from each subject after he had read each segment.

One version of the text was the "original" version of the published text taken directly from the technical manual initially prepared by the contractor. This text was prepared according to standard military specifications for technical manuals. The second version was a "revised" version subsequently prepared by the contractor in accordance with guidelines intended to simplify the readability of the text as assessed by standard readability equations. Generally, the guidelines directed the contractor to examine the prose for long or complex sentences, use of the passive voice, vague and ambiguous terms, and the use of unnecessary technical words. More specifically, the contractor was advised that the revised version should meet the following criteria: 1) an overall grade level of 10.0 as measured by the Kincaid Readability Formula; 2) no segment with a grade level exceeding 13.0 as measured by the Kincaid formula; 3) no sentence greater than 30 words in length; 4) all sentences in the active voice; 5) no sentence that begins with a preposition; and 6) all simple sentences with no more than two phrases. Samples of parallel segments of the original and revised versions of the text are shown in Table 1.

Experimental Procedures

All subjects participated in the study in individual sessions. They were told that they would be reading segments of text taken from the NATO Seasparrow Technical Manual and asked to tell how they would explain the

Table 1

Parallel Segments of the Original and Revised Text

Original Version

Air Designate Status. During this function, all COMPUTER COMPLEX DIGITAL COMPUTER functions are sequenced from the ready mode to the air designate mode status. The status change is reported to the digital and analog EDS by the EDS repeatback function as the DIGITAL and ANALOG EDS REPEATBACK DATA consisting of the RADAR A/B MODE STATUS signal. The air designate status is also reported over the 30-bit RSC A/B DOC DATA word which is routed via the SDC OUTPUT CHANNELS and the RADAR A/B RSC DOC TRANSFER CIRCUIT (A9). The decoded IND WORD 1 content is provided to the AIR MODE CIRCUIT as the AIR MODE signal and to the DESIGNATE MODE CIRCUIT as the DESIGNATE signal. The AIR MORE CIRCUIT produces an AIR MODE status signal which is distributed as the AIR MODE signal and provided to the FOC STATUS IND CIRCUIT where it is displayed. The RSC DESIGNATE MODE CIRCUIT produces a DESIGNATE MODE status signal which is distributed as the DESIGNATE MODE signals provided to the RADIATE CONT CIRCUIT of the radar radiate circuit function, and provided to the FOC STATUS IND CIRCUIT where it is displayed.

Revised Version

Air Designate Status. The air designate function sequences all digital computer functions from ready mode to air designate mode. This status change is reported to the digital and analog EDS. The RADAR A/B MODE STATUS signal (B3) is applied as DIGITAL EDS REPEATBACK DATA signals to the digital EDS. The RADAR A/B MODE STATUS signal is also applied as ANALOG EDS REPEATBACK DATA signals to the analog EDS via the SDC output channel circuitry. The air designate status is also reported to the RSC via the DOC transfer circuit (A9) as part of the RSC A/B DOC DATA signal (A4). The AIR MODE and DESIGNATE MODE signals are decoded from the IND WORD 1 signal (A9) and applied to the air and designate mode circuits respectively. The air mode circuit produces an AID MODE signal which is applied to the FOC status ind circuit where it is displayed. The AIR MODE signal is also distributed to logic circuitry within the RSC. The designate mode circuit (A10) produces a DESIGNATE MODE signal. The DESIGNATE MODE signal is applied to the radiate control circuit (B11), and to the FOC status indicator circuit where it is displayed on the FOC.

material to one of their technician colleagues. Subjects were asked to imagine someone they worked with who was just learning the system and actually picture themselves explaining the passages to him. Subjects were asked to identify the name of someone on their ship to whom they could imagine making these explanations.

Before beginning to read the text, subjects were provided a sheet of paper identifying four pieces of hardware on the NATO Seasparrow system. They were asked to draw a schematic diagram on the given sheet indicating the major hardware components and their interconnections.

After completing the schematic drawing, subjects were provided with one of the text versions and the 7-page schematic drawing from the technical manual. They were asked to read each segment and to use the schematic as they wished. After reading and studying each segment, subjects were asked to explain aloud what they had read to the "buddy" whom they had identified earlier. The sessions were tape-recorded for subsequent analysis and the experimenter recorded the time spent reading and studying, use of the schematic, and the time spent explaining. Finally, subjects were asked to make another schematic drawing of the system hardware and interconnections on a new sheet of paper (with the same identified hardware).

RESULTS

The results of the study can be organized around analyses of three

sources of data: the technical prose, transcripts of the verbal protocols of subjects audiotaped during the experiment, and the schematic drawings of subjects.

Three segments were sampled for analysis of the prose and the corresponding protocols. These segments were judged by two electrical engineering experts as low, moderate, and more difficult in technical complexity, respectively. Furthermore, they occurred after the warm-up phase of the experiment during which subjects were getting accustomed to the procedure of speaking aloud, but before fatigue set in.

The following sections present the results emerging from the analyses of the three primary data sources. Analyses of the prose are described first followed by a discussion of the protocol and schematic analyses.

Prose Analysis

First, the syntactic characteristics of the sampled prose segments were analyzed to determine the extent to which the guidelines were adhered to in the text revision. Second, the semantic characteristics of the two versions of text were examined to explore their comparability in meaning.

Syntactic Analysis. The syntax of the sampled segments was analyzed using the Writer's Workbench program developed by Bell Laboratories. This computer program provides information on a variety of syntactic indicators including readability level, sentence length, sentence types, verb usage, and sentence beginners.

Table 2 summarizes the results for the original and revised versions of the three sampled segments. These findings indicate that in general the revised text was simplified according to the guidelines provided to the

Table 2

Syntactic Characteristics of Original (O) and Revised (R) Text

Syntactic Indicator	Segment 4		5		7	
	O	R	O	R	O	R
Kincaid Readability Level	13.7	10.8	14.9	10.2	16.0	12.3
Average Sentence Length (in words)	27.0	19.5	27.3	17.1	29.3	21.0
Sentence Type:						
% Simple	43	60	36	67	50	89
% Complex	29	40	18	0	38	0
% Compound-Complex	29	0	27	7	0	0
% Compound	0	0	18	27	13	11
% Passive Verbs	56	67	48	48	55	60
% Sentences Beginning with a preposition	14	0	9	0	13	0

contractor. The readability levels of the original versions were reduced 3-4 grade levels in the revised versions as measured by the Kincaid readability formula. Sentence length was substantially reduced and the percentage of simple sentences was increased in the revised compared to the original versions. As directed, the contractor did modify sentences beginning with a preposition in the revised versions; however, the percentage of passive verbs was not systematically reduced.

In sum, the syntax of the revised version was simplified from the original version. However, passive verb constructions were frequent in both versions.

Semantic Analysis. The semantic content of the sampled segments of prose was analyzed using two sets of procedures. First, the electrical engineering experts identified the "gist" units or main ideas, and the information units represented by the content. In addition, they rated the information units on a number of dimensions, such as hierarchical level and type of topic. Second, a propositional analysis of the prose content was undertaken using the system developed by Kieras for use with technical text (Bovair and Kieras, 1981).

Turning first to the analyses of the subject matter experts, one of the electrical engineering experts formulated gist units to represent the macrostructure of the text and information units to represent its microstructure. The gist units were developed after Pearson and Camperell (1981). One gist unit was prepared for each segment of the original and revised text. Each gist unit consisted of a one sentence summary of the

main idea of the text segment. Thus, each gist unit provides an overview of the text segment and roughly approximates terms such as key episode, theme, main plot, and top-level discourse structure used by other writers.

Information units were devised using the guidelines and criteria of Van Matre (1975). Each information unit consisted of a simple sentence that expressed a single fact. In re-writing the text into information units, the electrical engineering expert (EEE) was guided by the following criteria:

1) Rewrite compound sentences as a group of simple sentences.

2) Reconstruct compound subjects, verbs, and objects as a group of simple sentences expressing a single fact per sentence.

3) Rewrite series of phrases as simple sentences that express a single fact per sentence.

4) Treat compound nouns representing a single term as a single word rather than as a noun preceded by a group of modifiers.

5) Use the active voice where possible.

Part A of Table 3 summarizes the results of the initial analyses of the subject matter expert. Each version of the sampled segments contained the same number of gist units. Further, the gist units were identical for each pair of segments suggesting that the revision did not alter the overall nature or amount of information presented.

The original and revised versions did differ in the number of information units that they contained. While both versions of segment 4 were equivalent on this dimension, the revised version of segment 5 contained more information units than the original version. This pattern was

Table 3

Semantic Analysis of Prose Using

Information and Gist Units

Semantic Indicator	4		5		7	
	O*	R*	O*	R*	O*	R*
A. Summary of Initial Analysis						
No. of Gist Units	3	3	5	5	2	2
No. of Information Units	24	24	26	34	27	20
No. of Sentences	6	10	10	15	8	9
"Density	4.0	2.4	2.6	2.3	3.4	2.2
B. Nature of Information Units						
Hierarchical Level**	2.11	1.83	1.96	1.70	2.30	1.75
Topic-% System	15.8	16.7	23.1	39.4	29.6	25.0
% Signal	21.1	29.2	11.5	18.2	18.5	5.0
% Hardware	21.1	0	38.5	12.1	11.1	10.0
% Signal + Hardware	42.0	54.1	19.2	30.3	22.2	40.0
% Signal + System	0	0	3.85	0	14.8	10.0
% System + Hardware	0	0	3.85	0	3.8	10.0
Salience*	1.79	1.33	1.42	1.52	1.93	1.30

* Where O = Original and R = Revised
** Rated on a 3-point scale where 1=high and 3=low.

Table 4

Propositional Analysis of Orignial (O) and Revised (R) Text
(after Kieras)

Measure	Segment 4				Segment 5				Segment 7			
	O		R		O		R		R		O	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Propositions per sentence	1.17	1.60	4.70	1.77	8.80	4.59	5.87	2.10	8.63	4.31	5.00	1.58
2. By type per sentence:												
verb-based	2.83	1.47	1.60	1.07	3.70	1.77	2.47	1.41	2.88	1.55	1.67	0.71
modifiers	1.83	1.83	1.60	1.58	2.60	2.88	1.93	1.44	3.38	2.07	1.44	1.13
prepositions	2.67	0.82	2.00	1.15	2.30	1.34	2.27	3.61	2.75	1.67	1.78	1.09
3. Embedded propositions per sentence	2.50	1.05	2.00	1.25	2.80	3.33	1.40	1.59	3.38	2.83	2.00	0.87
4. Propositions with missing subject per sentence	1.50	1.05	1.10	1.10	1.60	2.72	0.80	1.15	0.63	0.74	0.67	0.71

reversed for segment 7. However, the revised versions did contain more sentences than the original versions. Thus, when a measure of idea density was constructed (a ratio of the number of information units to the number of sentences), the density of the revised versions was lower than that of their respective original versions.

Subsequently, the second subject matter expert examined the nature of the information units by rating each one on hierarchical level and salience as well as identifying the type of topic presented. Part B of Table 3 summarizes the results of this analysis.

Generally, the information units in the revised versions tended to present ideas at a higher hierarchical level (i.e., main ideas rather than supporting details) and with greater salience to the overall passage. Further, information units in the revised versions tended to address the overall system or the interrelationship between the signal and the hardware to a greater extent than those in the corresponding original version.

Table 4 summarizes the results of the propositional analyses of the sampled segments. The propositional analyses constitute a more fine-grained examination of the text microstructure. As can be seen in Table 4, the revised versions contained fewer propositions per sentence than the original versions. Furthermore, there was less variability on this dimension in the revised versions. When the type of propositions per sentence was examined, (i.e., verb-based, modifiers, or prepositional), the revised versions tended to show a reduction in verb-based propositions (and to a lesser extent, modifiers) compared to the original versions. Further, when embedded propositions (propositions referring to other propositions within

the same sentence) were examined, their incidence was also reduced in the revised versions. Finally, the number of propositions with missing subjects, e.g., passive constructions, was slightly reduced in the revised versions in two of the sampled segments.

Thus, the results of both sets of analyses suggest that the original and revised versions of the sampled segments were equivalent in the overall amount and nature of the information provided. However, the sentences in the revised versions were simpler in their semantic content. They were lower in "idea density", and higher in hierarchical level and salience. Further, they tended to concern the overall system or the relationship between signal and hardware. The revised versions also showed fewer propositions per sentence and less embedding of propositions within sentences than the original versions.

Protocol Analysis

Subject's verbal protocols were used to investigate reader interaction with the text. These reader responses to the scenario of explaining the passage to a buddy were intended as an alternative to standard syntactic and semantic approaches to assessing readability. Procedures for scoring the protocols and obtaining reliability estimates for the scoring are described first. Then analyses of information unit summaries, gist unit summaries, and their relationship to individual differences among subjects are presented.

Scoring Rules and Reliability. Transcripts of the verbal protocols of subjects explaining aloud the material were used to examine subject's summaries of both gist and information units. Since the gist and

information units contained multiple propositions and partial summary was possible, rating scales were developed to assess the degree to which they were represented (rather than a binary yes/no judgment). Somewhat different procedures were used for scoring the gist and information units. A 5-point scale was used to score the gist units and scoring was conducted liberally based on the literal statements in the transcripts as well as inferences that followed from what they had said. During training, the two scorers were able to reach a high level of agreement (in fact 100%) using these procedures. (See Table 5.)

For the information units, a more stringent policy was adopted that required the same wording or synonyms in the transcripts. The two scorers were able to score the information units reliably using a 3-point scale where 1 = little or no inclusion, 2 = partial inclusion, and 3 = full or almost full inclusion. Their level of agreement was about 86% on the final training segment. (See Table 5.)

In addition, the scorers identified the frequency of connecting statements (with previous text segments and between the text and the schematic), elaborations beyond the text (system-specific or general electronics) and meta-statements about the audience or the summary task. During training, a 100% level of agreement was attained by the two scorers.

Table 5 also summarizes the reliability of scoring for the actual scoring. As shown in the table, high reliability was maintained on the information units and the other frequency counts. Interrater agreement fell on the gist units to 67%; however, disagreements were generally only one point off on a 5 point scale.

Table 5

Reliability Estimates for Protocol Scoring

Type of Unit	N	% Agreement	Mean Discrepancy Over all Items	Mean Discrepancy Over Items Where Disagreements
I. Actual Scoring				
Information Units (3-point scale)	155	90.87	0.12	1.05
Gist Units (5-point scale)	20	67.23	0.51	1.33
Other (Frequency Counts)	38	88.73	0.11	1.00
II. Final Training Segment				
Information Units (3-point scale)	28	85.71	0.81	1.25
Gist Units (5-point scale)	3	100.00	0.0	0.0
Other (Frequency Counts)	6	100.00	0.0	0.0

Summary of Information Units. Subjects' summaries of the information units contained in the text were examined using two strategies. First, the relationship of subject experience and text version (the two primary independent variables in the study) on the amount of summary of information units was examined using two-way analyses of variance. Second, the influence of these variables on the type of information units summarized was examined.

The results of the first set of analyses are summarized in Table 6. Separate analyses were conducted for each text segment with amount of summary measured by subjects' average summary ratings over all information units in the segment. The main effects for experience (E) and text (T) as well as their interaction (EXT) were statistically significant for all three sampled segments. When the cell means were graphed and examined using Scheffe's test for post hoc comparisons (Kirk, 1982), a similar pattern emerged in all three cases. The experienced non-NSS technicians showed significantly greater information unit summary with the revised compared to the original text. Further, with the revised text the experienced non-NSS technicians performed significantly above the experienced and novice NSS technicians. The three experience groups did not differ significantly with the original version nor did the experienced and novice NSS technicians differ from each other with the revised version.

The second set of analyses examined the type of information units summarized based on ratings and categorizations made by one of the subject matter experts. These included hierarchical level of the information unit, salience of the unit to the overall gist of the segment, and type of

Table 6

Influence of Experience and Text on Information Unit Summary (3-point Scale) of Original (O) and Revised (R) Text

Segment 4

a. ANOVA

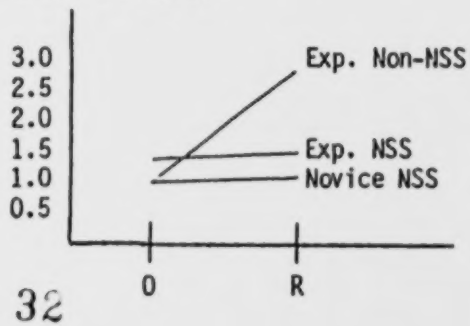
Source	df	SS	F
E	2	2.7089	17.23***
T	1	2.0842	26.53***
EXT	2	2.1082	13.41***
Error	12	0.9433	

***p<.001

b. Cell Means

	O	R	
Exp. Non-NSS	1.22	2.86	2.04
Exp. NSS	1.25	1.57	1.41
Novice NSS	1.07	1.15	1.11
	1.18	1.86	1.52

c. Graphic Display



Segment 5

a. ANOVA

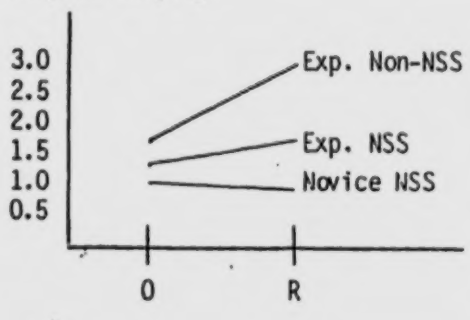
Source	df	SS	F
E	2	2.6099	22.93***
T	1	1.5744	27.66***
EXT	1	1.4349	12.61***
Error	12	0.6830	

***p<.001

b. Cell Means

	O	R	
Exp. Non-NSS	1.53	2.87	2.20
Exp. NSS	1.35	1.78	1.57
Novice NSS	1.29	1.28	1.29
	1.39	1.98	1.69

c. Graphic Display



Segment 7

a. ANOVA

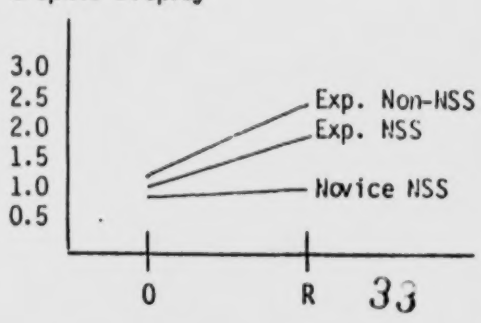
Source	df	SS	F
E	2	1.4726	13.35***
T	1	2.3835	43.22***
EXT	2	0.6627	6.01**
Error	12	0.6618	

**p<.01
***p<.001

b. Cell Means

	O	R	
Exp. Non-NSS	1.32	2.55	1.94
Exp. NSS	1.26	1.92	1.59
Novice NSS	1.09	1.38	1.23
	1.22	1.95	1.59

c. Graphic Display



topic. The first two types were measured on a 3-point rating scale where 1 = high and 3 = low, while the last was based on a category system of six possible topics or combinations of topics.

Separate analyses were conducted for each text segment. Each analysis included two between-subjects variables (experience and text) and one within-subjects variable (type of information unit). The dependent measures for each subject were average comprehension ratings for information units at each level or category of the scale being examined (e.g., hierarchical levels 1, 2, 3).

Table 7 summarizes the results for hierarchical level and Table 8 for topic. (No significant differences were obtained for the salience ratings, so the results are not presented here.)

As shown in Table 7, the between-subjects portion of the analyses replicated the effects of experience level, text version and their interaction described earlier. The within-subjects portion revealed a significant main effect for hierarchical level for segments 5 and 7. Post hoc comparisons among the marginal means using Scheffe's test indicated that subjects tended to show more summary of the units rated high in hierarchical level (i.e., main ideas) compared to those rated lower in hierarchical level (i.e., supporting details).

As displayed in Table 8, the effects of experience level, text version, and their interaction were again replicated as expected. In addition, statistically significant interactions between topic and text were obtained for segments 5 and 7. (A main effect for topic with segment 7 was also obtained.) Post-hoc comparisons of the means for the topic by text interaction were conducted using Tukey's HSD test (Kirk, 1982).

Table 7

Influence of Hierarchical Level on Information Unit Summary

Segment 4

a. ANOVA

Source	df	SS	F
E	2	6.2881	5.63**
T	1	6.8152	12.20**
ExT	2	5.2440	4.70**
Error	12	6.7014	
HL	2	1.2825	2.95
HLxE	4	1,8232	2.10
HLxT	2	0.3600	0.83
HLxExT	4	5.2196	
Error	24	5.2196	

** p < .01

b. Marginal Means for HL

HL1	1.81
HL2	1.44
HL3	1.66

Segment 5

a. ANOVA

Source	df	SS	F
E	2	8.7549	16.15**
T	1	4.0057	14.78***
ExT	2	5.2518	9.69**
Error	12	3.2528	
HL	2	0.7748	4.25*
HLxE	4	0.0633	0.17
HLxT	2	0.6057	3.32
HLxExT	4	0.2721	0.75
Error	24	2.1868	

* p < .05
** p < .01
*** p < .001

b. Marginal Means for HL

HL1	1.83
HL2	1.67
HL3	1.53

Segment 7

a. ANOVA

Source	df	SS	F
E	2	5.9826	19.29***
T	1	6.8603	44.23***
ExT	2	2.8030	0.94**
Error	12	1.8613	
HL	2	2.4432	9.46***
HLxE	4	0.9288	1.80
HLxT	3	0.4295	1.66
HLxExT	4	1.2110	2.35
Error	24	3.0979	

** p < .01
*** p < .001

b. Marginal Means for HL

HL1	1.95
HL2	1.48
HL3	1.53

Table 8

Influence of Topic on Information Unit Summary

Segment 4

a. ANOVA

Source	df	SS	F
E	2	5.8682	7.29**
T	1	5.8246	14.46**
EXT	2	5.4860	6.81**
Error	12	4.8328	
Top	2	1.1753	2.22
TopxE	4	1.1179	1.05
TopxT	2	0.3901	0.74
TopxExT	4	0.4449	0.42
Error	24	6.3601	

** p < .01

Segment 5

a. ANOVA

Source	df	SS	F
E	2	9.1671	16.62***
T	1	6.2143	22.53***
ExT	2	5.0573	9.17**
Error	12	3.3096	
Top	3	0.4039	1.43
TopxE	6	0.8806	1.55
TopxT	3	1.3642	4.82**
TopxExT	6	1.3800	2.44
Error	36	3.3993	

* p < 0.5
** p < .01
*** p < .001

b. Means for TopxT Interaction

	O	R
System	1.56	1.98
Signal	1.59	1.85
Hardware	1.32	2.31
Signal & Hardware	1.27	1.96

Segment 7

a. ANOVA

Source	df	SS	F
E	2	7.9206	21.867***
T	1	19.3661	107.06***
ExT	2	4.1333	4.42**
Error	12	2.1708	
Top	5	5.7686	5.90***
TopxE	10	2.8106	1.44
TopxT	5	4.6316	4.74***
TopxExT	10	1.1552	0.59
Error	60	11.7261	

* p < .05
** p < .01
*** p < .001

b. Means for TopxT Interaction

	O	R
System	1.46	2.07
Signal	1.16	1.56
Hardware	1.07	1.72
Signal & Hardware	1.00	1.68
System & Signal	1.33	2.39
System & Hardware	1.00	2.67

The pattern of results was somewhat different for segments 5 and 7. With segment 7, system-related topics tended to be summarized more than more specific topics, particularly with the revised version. With segment 5, there was greater summary of hardware with the revised version compared to the original text.

Summary of Gist Units. The influence of experience and text version on subjects' overall summary of gist units was also examined using two-way analyses of variance. Again, analyses were conducted separately for each segment where gist unit summary was measured by average gist unit summary ratings for each segment. A somewhat complicated pattern of results emerged, as shown in Table 9. For segment 5, a similar pattern of results occurred as described above for amount of information unit summary. Both experienced non-NSS and NSS technicians showed significantly greater gist unit summary with the revised compared to the original text version. The novice NSS technicians did not show significant differences in performance with the original and revised versions and they performed significantly below the two groups of experienced technicians on the revised version. For segment 7, all three experience groups showed greater gist summary with the revised compared to the original version. In addition, the two experienced groups (non-NSS and NSS) showed significantly greater gist comprehension than the novice NSS technicians. For segment 4, only a main effect for experience was obtained suggesting that the groups were ordered on degree of gist summary as follows: experienced non-NSS technicians, experienced NSS technicians, and novice NSS technicians.

Table 9

Influence of Experience and Text on Gist Unit Summary (5-point Scale)

Segment 4

a. ANOVA

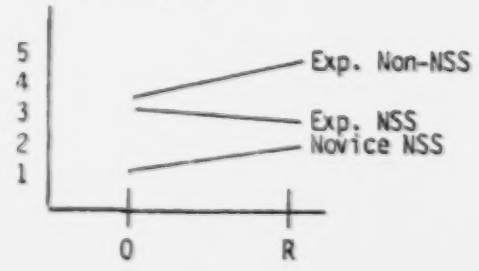
Source	df	SS	F
E	2	10.7160	3.91*
T	1	2.0000	1.46
EXT	2	2.3704	0.86
Error	12	16.4444	

*p<.05

b. Cell Means

	O	R	
Exp. Non-NSS	3.22	4.77	4.00
Exp. NSS	3.11	2.89	3.00
Novice NSS	1.78	2.44	2.11
	2.70	3.37	3.04

c. Graphic Display



Segment 5

a. ANOVA

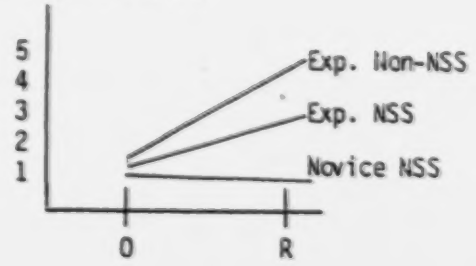
Source	df	SS	F
E	2	10.0844	9.22**
T	1	7.2200	13.21**
EXT	2	8.0133	7.33**
Error	12	6.5600	

**p<.01

b. Cell Means

	O	R	
Exp. Non-NSS	1.87	4.80	3.33
Exp. NSS	1.80	3.00	2.40
Novice NSS	1.67	1.33	1.50
	1.78	3.04	2.41

c. Graphic Display



Segment 7

a. ANOVA

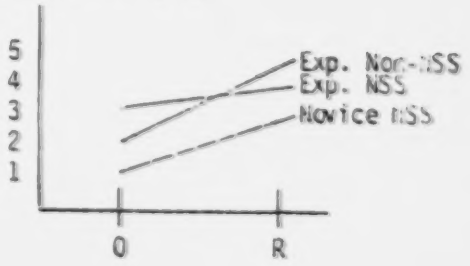
Source	df	SS	F
E	2	7.1111	4.13*
T	1	7.3472	8.53**
EXP	2	1.4444	0.84
Error	12	10.3333	

*p<.05
**p<.01

b. Cell Means

	O	R	
Exp. Non-NSS	3.00	4.83	3.92
Exp. NSS	3.67	4.17	3.92
Novice NSS	1.83	3.33	2.58
	2.83	4.11	3.47

c. Graphic Display



BEST COPY AVAILABLE

A similar analysis was conducted for the frequency of connectives, elaborations, and meta-statements. No statistically significant effects were isolated for experience or text. In fact, the frequency of such statements was generally quite low. The most frequent type of statement was system-specific elaborations which averaged 1.39 per segment (s = 2.89).

Relation to other Individual Difference Measures. The relationship of information and gist unit summary to time spent reading and studying, use of the provided schematic diagrams, and individual differences among subjects as measured by their Armed Forces Qualification Test scores (AFQT) and their Armed Services Vocational Aptitude Battery (ASVAB) scores was examined. No significant relationships were identified.

In sum, the revised text version contributed to greater summary of information units for the experienced non-NSS and NSS technicians but did not appear to facilitate the information unit summary of the novice NSS technicians. There was also some evidence that subjects tended to comprehend more global ideas rather than supporting details, particularly with the revised text.

Schematic Analysis

The schematic drawings prepared by subjects before and after reading the text were scored using a system developed by one of the subject matter experts. The scoring system produced two sub-scores and a total score. The first sub-score was concerned with the functional topology or nodes of the drawing, e.g., the functional boxes such as the Computer Complex. The nodal hierarchies were additive and weighted in a geometric fashion by

BEST COPY AVAILABLE

BEST COPY AVAILABLE

level of specificity to emphasize the level of knowledge. Thus, a person could earn a score of 1, 2, 4, 8, or 16 for the computer complex, for example, depending on the level of detail. The second sub-score concerned the signal flow or connections between nodes. One point was assigned for a connection between boxes, two points for the direction of the path, and four points if the connection was named. The scoring system also contained a penalty for incorrect nodes and signals such that points were subtracted for errors drawn. The total score was simply the sum of the function and signal sub-scores.

Because the sub-score distributions were highly skewed, the first step in the analysis was to truncate the distributions and to transform the original scores to a five-point scale. This transformation was accomplished by examining the frequency distributions of the sub-scores and dividing them at naturally occurring breaks between score values.

Analysis of the pre-test drawings indicated a significant main effect for experience on both the function sub-score and the signal sub-score (see Table 10). Post-hoc comparisons between means using Scheffe's method revealed that, on the function sub-score, the experienced and novice NSS technicians performed significantly above the experienced non-NSS technicians. On the signal sub-score, the experienced NSS technicians showed significantly higher performance than either of the other two groups.

Table 11 presents an analysis of the posttest schematic drawings by experience and text version. There was a significant main effect for experience on the function sub-scores. Post-hoc comparisons among means

Table 10

Relationship of Experience and Text
To Pretest Schematic Drawings

A. Function Sub-Score

1. ANOVA

Source	df	SS	F
E	2	12.6627	8.19**
T	1	3.1405	4.06
EXT	2	1.2261	0.79
Error	11	8.5000	

** $p < .01$

2. Experience Group Means

Exp. Non-NSS	2.17
Exp. NSS	4.17
Novice NSS	3.60

B. Signal Sub-Score

1. ANOVA

Source	df	SS	F
E	2	9.3000	5.48**
T	1	0.9444	1.11
EXT	2	4.4333	2.61
Error	11	9.3333	

** $p < .01$

2. Experience Group Means

Exp. Non-NSS	1.50
Exp. NSS	3.00
Novice NSS	1.40

Table 11

Relationship of Experience and Text
to Post-test Schematic Drawings

A. Function Sub-Score

1. ANOVA

Source	df	SS	F
E	2	9.9353	4.15*
T	1	0.1381	0.12
EXT	2	0.9953	0.42
Error	11	13.1667	

*p<.05

2. Experience Group

	Means
Exp. Non-NSS	2.50
Exp. NSS	4.00
Novice NSS	4.20

B. Signal Sub-Score

1. ANOVA

Source	df	SS	F
E	2	10.2510	3.28*
T	1	2.5621	1.64
EXT	2	4.1379	1.33
Error	11	17.1667	

*p<.07

2. Experience Group

	Means
Exp. Non-NSS	2.33
Exp. NSS	3.33
Novice NSS	1.40

using Scheffe's method showed that, as in the analyses of the pretest drawings, both the experienced and novice NSS technicians performed higher on this dimension than the non-NSS technicians.

The main effect for experience on the signal sub-scores approached statistical significance ($p<.07$); and the pattern of means was somewhat different than at the pretest. In this case, the experienced NSS technicians performed significantly higher than the novice NSS technicians as at pretest. However, the performance of the experienced non-NSS technicians was not significantly different from the experienced NSS technicians.

Finally, pretest and posttest scores were compared for each of the three experience groups using t-tests for dependent groups. While no statistically significant differences were observed, examination of the means suggests that the novice NSS technicians tended to show more improvement on the function sub-score from pretest to posttest, while the experienced non-NSS technicians tended to show improvement on the signal sub-score.

In sum, the NSS technicians exhibited greater understanding of the functional hardware in their schematic drawings than the non-NSS technicians, as would be expected based on their direct experience with the NSS system. Further, experienced NSS technicians showed a greater understanding of the signal paths at pretest than the other two groups. However, the experienced non-NSS technicians improved at posttest such that their signal performance was not significantly different from the experienced NSS technicians.

DISCUSSION

This research examined the potential use of scenarios simulating a real world task to assess the readability of general technical text. The scenario consisted of asking subjects to read a segment of text and to explain to a colleague how the particular component described in the passage functioned. Subjects' summaries were assessed based on the extent to which the gist of the passage was represented in their summaries as well as specific information units.

Two experimental variations were conducted to examine their effects on subjects' summaries in response to the scenario. The first variation involved the text with an original version and a revised version that was simplified using traditional readibility indices. The second variation centered on the world knowledge that subjects brought to the task with three groups of subjects differing in their amount and type of experience. Finally, subjects' graphic representations of their system as indicated by schematic drawings were examined to determine the influence of the text and subjects' world knowledge on functional understandings of the system.

Results supported the initial distinctions drawn between text versions and experience groups. The revised version was indeed syntactically simpler on a number of dimensions such as readability level, sentence length, and sentence structure. Furthermore, while the overall semantic content or gist of corresponding segments remained comparable, the revised version was lower in density of ideas and ideas presented were higher in hierarchical level and salience than the original version.

Further, analyses for the pretest schematic drawings supported that there were pre-existing differences among the three groups of technicians. NSS technicians showed greater functional knowledge in their drawings than non-NSS technicians, while experienced NSS technicians showed greater signal knowledge than either of the other two groups.

Analyses of the verbal summaries suggest that the revised version had the greatest facilitating effects for experienced non-NSS technicans. Technicians in this group showed performance superior to both experienced and novice NSS technicians in analyses of both the amount and types of information summarized.

One might speculate that the simpler revised version of the text allowed the experienced non-NSS technicians to apply their general world knowledge about electronic systems and technical manuals in order to extract the information necessary to understand the specific operations of the Seasparrow System. Such an explanation aligns with work by Rumelhart and Ortony (1977) which suggests that efficient learners call up schema corresponding to the particular type of discourse and instantiate the schema based on what is read. The experienced non-NSS technicians were familiar with technical manuals generally, thus, they may have had a well-developed text scheme for this type of technical manual. Furthermore, they were familiar with electronic systems and may have had a content schema as well. The simplifications in the revised version may have facilitated the ability of the experienced non-NSS technicians to instantiate their pre-existing content schema with specific facts related to the Seasparrow System.

One might further speculate that the hands-on experience of the experienced NSS technicians actually interfered with their instantiation of facts from the text into their content schema. In several cases, members of this group noted that the text was incomplete or inaccurate (e.g., "it doesn't really work that way"). Meyer (1984) reported a similar phenomenon in which readers disagreed with an author's message, rejected his schema, and substituted their own schema with instantiations based on information outside the text.

The analyses of the verbal protocols also suggested that main ideas rather than supporting details tended to be summarized. This finding replicates the work of Meyer (1975) and others who have shown that height in the content structure is the best predictor of information recalled.

Analyses of the posttest schematic drawings revealed the same effects for experience as the pretest drawings. Changes in performance from pretest to posttest suggest that technicians with different types and amounts of experience may be extracting different types of information from the text. Novice NSS technicians tended to show improved functional understanding of the system while technicians experienced in systems other than the Seasparrow System showed improvements in signal knowledge.

The study reported here was exploratory and the results should be cautiously interpreted. However, there are indications that the use of scenarios and verbal summaries from subjects may provide valuable insights into the readability of text. Some advantages of the approach are that it can be used to establish a task environment that more closely parallels reading demands in the real world than much of the prior readability

research, that it views the reader as an active processor of text extracting meaning and incorporating meaning into preexisting cognitive structures, and that it yields rich data on what and how subjects use information contained in text.

Additional research is needed to clarify and operationalize the concept of readability. However, as this work and other research suggests, emerging concepts of readability must take into account the interaction of a reader (preexisting world knowledge, content and text schemas, goals for reading, for example) and the text (its syntactic and semantic characteristics). The use of task scenarios and verbal protocols may provide one means of learning more about this interaction so that guidance can be developed to make specific types of text more "readable" for particular types of readers.

Methodologically, the study illustrates a larger connection between the analytic schemes for assessing readability and for emerging procedures for assessing written productions. Although the relationship between reading and writing ability has received popular support of historical proportions, the demonstration of such a relationship depends upon the adaptation of common frameworks and comparable procedures for analyzing text. The cognitive vs. the communicative uses of writing suggest that writing is a way to clarify and then to demonstrate the clarity of conceptual understandings as evidence of the comprehensibility of written text. Present applied development in writing-from-reading (Quellmalz, 1984) demonstrates these interrelationships.

Yet, writing assessment itself has grown (as have concepts of readability) more sophisticated and has moved toward functional organization and semantic analysis. Newer scoring processes for writing depend less on the convention of syntax and more on what meaning and purpose is inherent in the text. Relatively simple procedures and rigorous training programs are available to teach the assessment of writing (Quellmalz & Burry, 1983) reliably. These are presently being applied in the ongoing international comparison of written composition supported by the IEA (Baker, 1982) and may well contribute to the task of protocol analysis. The analysis of oral protocols as indicators of the readability of text will undoubtedly remain of research rather than applied interest in education and training settings (until computerized speech analysis is perfected). Yet, scholars interested in either oral or written text analysis must begin to profit from the conceptual framework and research processes used by each other. Parallel effort need not go unacknowledged, and the analysis of language understanding can only benefit.

BEST COPY AVAILABLE

REFERENCES

Amirin, M.R. and Jones, B.F. Toward a new definition of readability. Educational Psychologist, 1982, 17, 13-30

Baker, E. The specification of writing tasks, Evaluation in Education, 1982, 5, 291-295.

Bovair, S., & Kieras, D.E. A guide to propositional analysis for research on technical prose. Technical Report #8. San Diego, CA: NPRDC, 1981.

Carver, R.P. Measuring the reading difficulty levels of Navy Training manuals (Tech. Rep.) Washington, D.C.: Office of Naval Research, September, 1974. (NTIS No. AD 780 448/7).

Bormouth, J.R. Development of readability analyses. (Contract #OEG-3-7-070053-0326). Washington D.C.: U.S. Office of Education, 1969.

Bransford, J.D. and Johnson, M.K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Duffy, T.M. Readability formulas: What's the use? In T. Duffy and R. Waller (Eds.) Designing Usable Texts. New York: Academic Press, 1985.

Duffy, T.M. Preparing Technical manuals: Specifications and guidelines. In D. Jonassen (Ed.) The technology of text. Vol II. Englewood Cliffs, N.J.: Educational Technology Publications, 1985b.

Duffy, T.M. Literacy instruction in the Armed Forces. Armed Forces and Society. In Press.

Duffy T.M., Curran, T.and Sass, D. Document design for technical job tasks: An evaluation. Human Factors, 1983, 25, 153-160.

Duffy T.M., & Kabance, P. Testing a readable writing approach to text revision. Journal of Educational Psychology, 1982, 74, 733-748.

Flesch, R.F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Flesch, R.F. The art of readable writing. New York: Harper, 1949.

Frase, L. Purpose in reading. In J.T. Guthrie (Ed.) Cognition, curriculum, and composition. Delaware: International Reading Assoc., 1977.

Kieras, D. Thematic processes in the comprehension of technical prose. In B. Britton and J. Black (Eds.) Understanding expository prose. England, N.J.: L. Erlbaum and Assoc., 1985a.

BEST COPY AVAILABLE

Kieras, D.E., Tibbits, H., Bovair, S. How experts and nonexperts operate electronic equipment from instructions. Technical Report #14. Arlington, VA: Office of Naval Research, 1984.

Kirk, R.E. Experimental Design, 2nd edition. Belmont, CA: Brocks/Cole Publishing Co., 1982.

Klare, G.R. The measurement of readabiliy. Ames, Iowa: Iowa State University Press, 1963.

Klare, G.R. A manual for readable writing. Glen Burnie, Maryland: REM Company (119A Roesler Road), 1975.

Klare, G.R. Readability standards for Army-wide publications. (Evaluation Report 79-1). Fort Benjamin Harrison, Ind.: U.S. Army Administrative Center, 1979.

Meyer, B.J.F. Text dimension and cognitive processing. In H. Mandl, N. Stein, & Trabosso, T.(Eds.), Learning and Understanding Texts.

Meyer, B.J. and Rice, G.E. Interaction of text variables and processing strategies for young, middle-aged, and older expert readers. Research Report #12. Tempe, AZ: Dept. of Educational Psychology, Arizona State Univ. 1983.

Miller, J.R. and Kintsch, W. Readability and recall of short prose passages: A theoretical analysis. Journal of experimental psychology: Human learning and memory., 1980, 6, 335-354.

Osborne, R. & Wittrock, M. Learning Science: A generative process. Science Education, 1983, 67, 489-508.

Pearson, P.D. The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. Reading Research Quarterly, 1974-1975, 10, 155-192.

Pearson, P.D., & Camperell, K. Comprehension of text structures. In J.T. Guthrie (Ed.), Comprehension and teaching: Research reviews. Newark, Delaware: International Reading Association, 1981.

Pichert, J.W. & Anderson, J.C. Taking different perspectives on a story. Journal of Educational Psychology, 1977, 69, 309-315.

Quellmalz, E.S. and Burry, J. Analytic scales for assessing students' expository and narrative writing skills. CSE Resource Paper No. 5, 1983.

Quellmalz, E.S. A system for analyzing functional writing: An example of research in service of policy and instruction. Paper presented at the Annual Meeting of the American Educational Research Associationm, New Orleans, 1984.

Redish, J.C., Felker, D., & Rose, A. Evaluating the effects of document design principles. Information Design Journal, 1982, 2/3/4, 236-243.

BEST COPY AVAILABLE

Rumelhart, D. & Ortony, A. The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & Montague, W.E. (Eds.), Schooling and Aquisition of Knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Spiro, R., Bruce, B., & Brewer, W. (Eds.) Theoretical issues in reading comprehension. Englewood, N.J.: L. Erlbaum and Assoc. 1980.

Swaney, J., Janik, C., Bond, S., & Hayes, J.R. Editing for comprehension: Improving the process through reading protocols. DUP-TR #14. Pittsburgh, PA.: The Communications Design Center, Carnegie-Mellon University. June 1981.

Van Matre, N.H., et al. Learning from lecture: Investigation of study strategies involving note-taking. San Diego, CA: NPRDC, 1975.

Wittrock, M. Generative reading comprehension. West Los Angeles, CA: University of California, 1982.

Wright, P. Editing: Policies and processes. In T. Duffy and R. Waller (Eds.) Designing Usable Texts. New York: Academic Press, 1985.

BEST COPY AVAILABLE